



OPEN

Association of GATA3 expression in triple-positive breast cancer with overall survival and immune cell infiltration

Xiuwen Chen^{1,2}, Weilin Zhao^{1,2}, Yugang Huang¹, Senyuan Luo¹, Xianbin Tang^{1✉} & Qiong Yi^{1✉}

Breast cancer remains a leading cause of cancer-related mortality among women, with triple-positive breast cancer (TPBC) being a particularly aggressive subtype. GATA binding protein 3 (GATA3) plays a crucial role in the luminal differentiation of breast epithelium and T-cell differentiation. However, the relationship between GATA3 and immune infiltration in TPBC remains unclear. This study collected and analyzed TPBC data from The Cancer Genome Atlas (TCGA), METABRIC, and GSE123845 databases. Univariate and multivariate Cox regression analyses, along with Kaplan–Meier survival analyses, were employed to assess the prognostic value of GATA3 and other clinical features. Subsequently, Gene Set Enrichment Analysis (GSEA) was conducted to explore the potential biological functions and regulatory mechanisms of GATA3 in TPBC. Additionally, ssGSEA analysis revealed the connection between GATA3 and immune infiltration. And the effects of neoadjuvant chemotherapy and immunotherapy on GATA3 expression were also explored. Finally, clinical samples were used to detect the relationship between GATA3 expression and tumor infiltrating lymphocyte (TIL) levels. Our results demonstrated that GATA3 was significantly overexpressed in TPBC tissues compared to normal tissues ($P < 0.05$). A positive correlation between GATA3 mRNA and protein levels was observed ($R = 0.55$, $P < 0.05$). Notably, high GATA3 expression was associated with poor overall survival (HR = 1.24, 95% confidence interval (CI) 1.25–11.76, $P < 0.05$). GSEA indicated significant enrichment of immune-related gene sets in low GATA3 expression groups. Furthermore, pathologic complete response (pCR) patients exhibited significantly lower GATA3 expression compared to residual disease (RD) patients. Mutation analysis revealed higher PIK3CA and TP53 mutation rates in high GATA3 expression groups. Finally, clinical validation data showed that the degree of TILs was significantly higher in the low GATA3 expression group. In conclusion, this study suggests that high GATA3 expression may be associated with poor prognosis and may reduce immune infiltration in TPBC.

Keywords GATA3, Triple-positive breast cancer, Tumor-infiltrating lymphocytes, Tumor microenvironment, Neoadjuvant chemotherapy

Breast cancer is a heterogeneous disease categorized into four distinct subtypes (Luminal A, Luminal B, HER2-enriched, and Basal-like) based on microarray-based gene expression profiling^{1,2}. Within the HER2-enriched breast cancer subtype, approximately 50% of tumors concurrently overexpress estrogen receptor (ER) and progesterone receptor (PR) in addition to HER2 amplification, suggestively named “triple positive” Breast cancer (TPBC)³. Compared to other subtypes, TPBC presents complex clinical features and prognoses. Some studies have indicated that TPBC is associated with higher tumor grades, significant lymph node metastases, and aggressive tumor behavior⁴. Conversely, others suggest that TPBC tends to have a better prognosis compared to hormone receptor (HR)-negative/HER2-positive breast cancer⁵. Although the combination of targeted therapy and endocrine therapy is generally considered effective for TPBC, the complex interactions between HER2 and the HR signaling pathway lead to increased tumor drug resistance and progression, necessitating further investigation into TPBC treatment⁶. Currently, neoadjuvant chemotherapy (NACT) is increasingly used for TPBC patients; however, few studies have reported biomarkers that predict TPBC prognosis and NACT efficacy. Additionally,

¹Department of Pathology and Department of Hematology, Taihe Hospital, Hubei University of Medicine, Hubei, China. ²These authors contributed equally: Xiuwen Chen and Weilin Zhao. ✉email: tangxianbin@yeah.net; yiqiong0428@163.com; tangxianbin@yeah.net

with recent advancements in immunotherapy for luminal-type breast cancer, the application of immunotherapy in breast cancer could be expanded⁷. Therefore, it is necessary to identify new biomarkers to better guide clinical assessment and treatment decisions beyond traditional perspectives in TPBC.

GATA3 is a transcription factor that plays a critical role in developing luminal differentiation of breast epithelium⁸ and T-cell differentiation⁹. Loss of GATA3 expression is associated with poor outcomes in previous clinical study¹⁰. Therefore, the immunohistochemical expression of GATA3 is widely used in clinical practice to differentiate breast metastases of unknown origin and assess prognosis, although there is a lack of reports on its prognostic significance in TPBC. Notably, GATA3 expression is strongly correlated with hormone receptors, with several studies suggesting a potential association with an improved response to endocrine therapy¹¹. Despite these established associations, the specific role of GATA3 in TPBC, particularly in relation to NACT response and immune infiltration, remains underexplored. Recent research has shown that the levels of TILs could predict the response to NACT in all molecular subtypes of breast cancer¹². However, research into the relationship between GATA3 expression and immune cell infiltration has been predominantly focused on bladder cancer¹³. Given the potential interaction between GATA3 expression and immune infiltration, investigating the relationship between GATA3 and TILs in TPBC could provide valuable insights into the mechanisms underlying treatment response and identify novel therapeutic targets.

In this study, we aimed to investigate the expression of GATA3 in TPBC and its association with immune infiltration. Initially, we examined GATA3 mRNA expression levels and conducted survival analysis in TPBC. Subsequently, the study provided further insight into the association between GATA3 expression, immune-related genes, and the immune infiltration of the tumor microenvironment (TME) through GSEA analysis. Additionally, we assessed the relationship between GATA3 expression and different NACT efficacy groups. Finally, relevant clinical cases were collected, and immunohistochemistry was employed to compare the levels of TILs between different GATA3 expression subgroups.

Materials and methods

Data collection and preprocessing

Transcript Per Million (TPM) normalized expression profile data and breast cancer sample information were obtained from The Cancer Genome Atlas (TCGA) database using the TCGABlinks R package^{14–16}. TPBC patients were identified based on ER and PR immunohistochemical results, and HER2 results derived from the clinical information, with the detailed screening process depicted in Fig. 1. Ultimately, 109 TPBC cases were included in this study. ER/PR positive status was established based on positive immunohistochemistry (IHC) results,

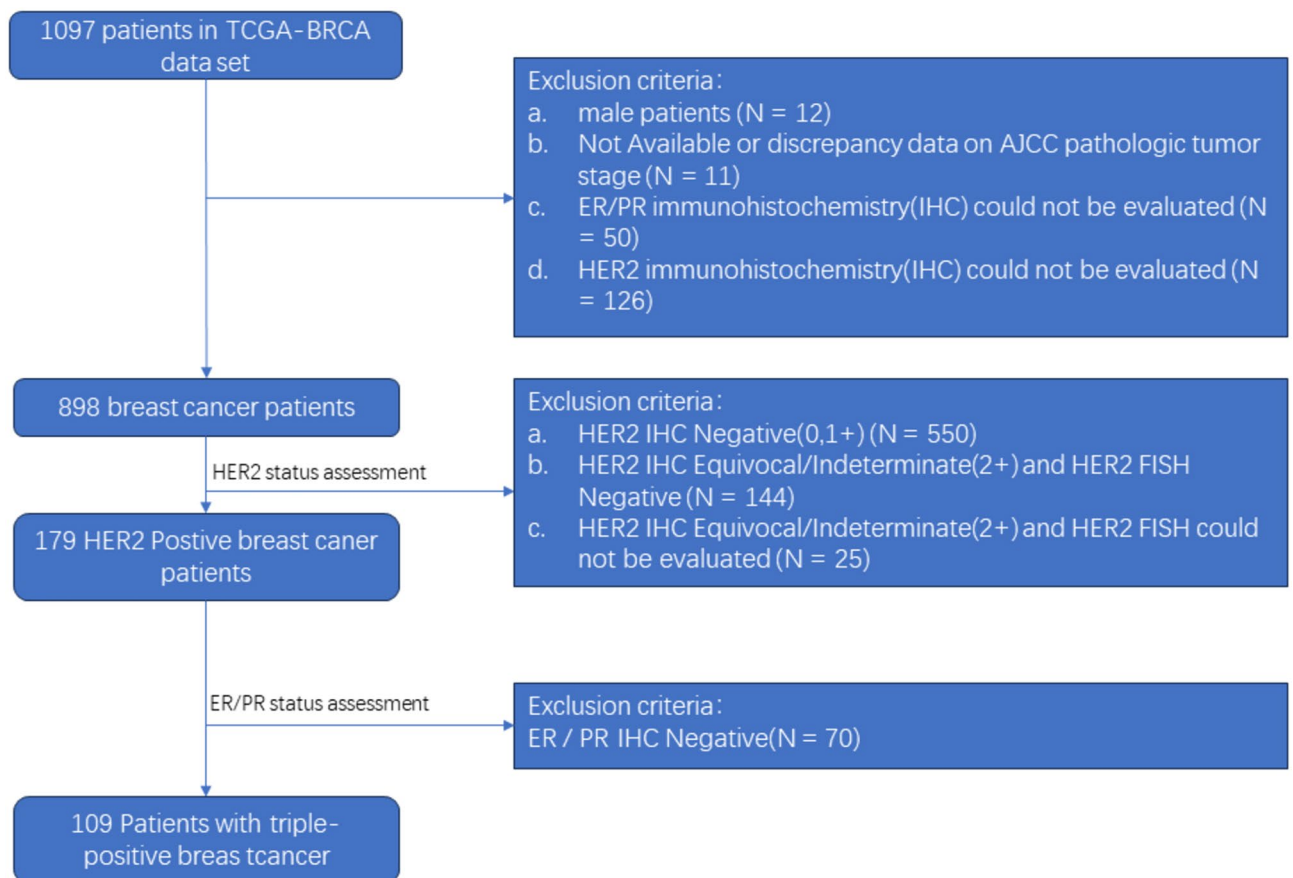


Figure 1. Screening process of TPBC cases in TCGA database.

and HER2 status was primarily determined according to the American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) HER2 testing guideline¹⁷. The METABRIC dataset, chosen for validation from the cBioPortal database (<https://www.cbioportal.org/>), comprised 46 TPBC cases identified based on ER/PR and HER2 status from the corresponding clinical information. To test whether GATA3 expression correlated with the response to NACT, the transcriptome data from TPBC patients in the public dataset GSE123845 was chosen¹⁸. The baseline clinicopathological characteristics of the patients in the different cohorts are summarized in Table S1.

Expression and prognostic significance of GATA3 in TPBC

Initially, GATA3 mRNA and protein expression in normal and TPBC tissues were initially analyzed. The protein expression data were derived from Reverse Phase Protein Array (RPPA) technology, a high-throughput proteomic technology used to assess protein expression levels in many samples simultaneously. Subsequently, TPBC patients were divided into high and low-expression groups based on the optimal cutoff point of GATA3 expression levels calculated by the survminer package. This package was employed to examine the survival data and find a cutoff point that effectively distinguishes between high-risk and low-risk groups according to their survival outcomes. Kaplan–Meier analysis was used to compare the impact of GATA3 expression on the overall survival of TPBC patients. Univariate regression analysis was initially conducted, and variables with statistically significant differences were subsequently included in the multivariate regression analysis. Finally, the correlation between GATA3 expression and various clinical characteristics, including age, Ki67 status, and histological grade, was investigated.

Gene set enrichment analysis

To explore the potential regulatory mechanisms of GATA3 in TPBC, Gene Set Enrichment Analysis (GSEA) was performed using the ClusterProfiler R package¹⁹. The analysis was conducted using annotated gene sets, including “h.all.v2023.1.Hs.entrez.gmt (Hallmarkers)”, “c2.all.v2023.1.Hs.entrez.gmt”, and “c5.all.v2023.1.Hs.entrez.gmt”. The statistical threshold was set at a false discovery rate (FDR) of < 0.05. Details on the subsequent analysis of immune infiltration related to the enrichment of immune-related gene sets are provided in the next section. To assess the enrichment of mismatch repair (MMR)-related gene sets, additional analyses were conducted on tumor mutation burden (TMB) and microsatellite instability (MSI). Relevant TMB and MSI data were obtained from the cBioPortal database (<http://www.cbioportal.org/>).

Association between GATA3 expression and immune infiltration

Immune-related genes were sourced from the ImmPort data portal (<https://www.immport.org/home>), and 2,483 immune-related genes were obtained. Given the large number of immune genes, the analysis of individual genes alone may not effectively assess the comprehensive biological processes and pathways associated with immune gene alterations. Therefore, Gene Set Variation Analysis (GSVA) was employed using 17 immune gene sets (Table S2). The single-sample Gene Set Enrichment Analysis (ssGSEA) method, implemented in the GSVA R package²⁰, was used to estimate the ratio of immune cell infiltration between the GATA3 high- and low-expression subgroups. The results were visualized through boxplots generated by the ggpubr R package. To evaluate the stromal and immune components and the overall tumor purity, the stromal, immune, and estimate scores were calculated using the estimate R package²¹. A statistically significant threshold was set at $P < 0.05$. The Cancer Immunome Atlas (<https://tcia.at/>) is a valuable resource for the comprehensive analysis of intratumoral immune landscapes and cancer antigenomes for 20 distinct solid cancers. We extracted the Immune Phenotype Score (IPS) data from TPBC patients for subsequent analysis. These scores could predict the response to immunotherapy, including anti-PD-1 and anti-CTLA-4 treatments²². The oncoPredict R package was used to predict drug sensitivity in different GATA3 expression groups²³.

Association between GATA3 expression and NACT response

To investigate the correlation between GATA3 expression and the efficacy of NACT, TPBC cases in the GSE123845 dataset were selected. In this dataset, patients were divided into pathological complete response (pCR) and residual disease (RD) groups. First, we compared the expression of GATA3 between RD and pCR patients and then used ssGSEA analysis further to analyze the differences in tumor-infiltrating immune cells. Finally, the potential association between GATA3 expression and the infiltration level of immune cells was explored.

Mutation and methylation analysis

GATA3 mutation and epigenomic (methylation) analyses in TPBC were performed using the cBioPortal for TCGA (<http://www.cbioportal.org/>) platform. This study compared mutational signatures across different GATA3 expression groups and examined differences in GATA3 expression among various GATA3 mutation groups in TPBC. Furthermore, the correlation between GATA3 expression and methylation levels was explored. Spearman's correlation coefficients were used to calculate the association between GATA3 expression and methylation.

Validation by immunohistochemistry staining in clinical specimens

Thirty-one formalin-fixed paraffin-embedded tissue samples were collected from patients diagnosed with primary TPBC at Taihe Hospital. Immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) methods were utilized to evaluate ER, PR, and HER2 expression status according to the definitions provided by the ASCO/CAP¹⁷. ER and PR status was based on tumor nuclei staining, with a positivity threshold of $\geq 1\%$. HER2 positivity was defined as an IHC score of 3+. For cases with an IHC score of 2+, FISH analysis was subsequently

conducted, and only patients demonstrating HER2 gene amplification were categorized as HER2 positive. The GATA3 status of all primary tumors was assessed through IHC staining using standard procedures (GATA3 antibody: L50-823, Maxim, China). Positive expression of GATA3 was observed in the nucleus of tumor cells. Each slide was randomly selected for analysis in ten high-power fields (200×). The semi-quantitative combined H-score was used to assess the intensity of GATA3 expression. The H-score was calculated by multiplying the staining intensity and the percentage of immunoreactive cells, with staining intensity graded on a scale of 0 (negative), 1 (weak), 2 (moderate), 3 (strong), and the percentage of immunoreactive cells ranging from 0 to 100%²⁴. This study categorized samples with an H-score ≥ 200 as the high-expression group, while the remaining samples were classified into the low-expression group¹⁰. Tumor-infiltrating lymphocytes (TILs) in the primary tumor were assessed with hematoxylin and eosin (HE)-stained slides and evaluated according to the International Immuno-Oncology Biomarker Working Group guidelines²⁵. The TILs score is determined by the percentage of tumor stroma area occupied by lymphocytes.

Statistical analysis

All statistical analyses were performed using R software version 4.1.2. A t-test was used to compare data between the two groups for continuous variables with a normal distribution. When continuous variables did not follow a normal distribution, the Wilcoxon rank-sum test was utilized for comparisons. Categorical data analysis was conducted using the chi-square or Fisher's exact tests. The Kaplan–Meier method was used to conduct survival analysis. A $P < 0.05$ was considered statistically significant.

Institutional review board statement

This study was approved by the Ethics Committee of Shiyan Taihe Hospital, Hubei University of Medicine. All experiments were performed according to regulations. Informed consent was obtained from patients in the study.

Results

The expression levels of GATA3 and the correlation between ER and HER2 expression in TPBC

To investigate the fundamental expression patterns of GATA3, mRNA expression levels were examined in normal and TPBC samples from the TCGA database. The analysis revealed a significant elevation of GATA3 expression in TPBC at the transcriptome level (Fig. 2A). Statistical analysis showed significant upregulation of GATA3 in TPBC compared to normal samples ($P < 0.05$). Additionally, Fig. 2B illustrated a positive correlation between GATA3 mRNA and protein levels (Spearman: $R = 0.55$, $P < 0.05$). A negative correlation was observed between ESR1 (ER) and ERBB2 (HER2) expression in TPBC (Spearman: $R = -0.28$, $P < 0.05$) (Fig. 2C).

The prognostic value of GATA3 expression in TPBC

Kaplan–Meier survival analysis was initially performed to investigate the correlation between GATA3 expression levels and overall survival, indicating that GATA3 was a significant unfavorable prognostic marker in TPBC ($P < 0.05$) (Fig. 3A). These findings were further validated in an independent METABRIC dataset (Fig. 3B). Additionally, univariate and multivariate Cox regression analyses were conducted to evaluate overall survival in TPBC patients. The results revealed that elevated GATA3 expression was significantly correlated with an unfavorable prognosis in TPBC (HR = 1.24, 95% CI 1.25–11.76, $P < 0.05$) (Fig. 3C). However, multivariate Cox regression analysis results indicated that GATA3 is not an independent prognostic factor for TPBC patients when considering age and clinical stage (Fig. 3D). Although GATA3 gene expression showed no differences across tumor stages in TPBC (Fig. 3E), subsequent analysis indicated a positive correlation between GATA3 gene expression and the T stage in TPBC (Fig. 3F) while showing no relation to the N stage (Fig. 3G). Finally,

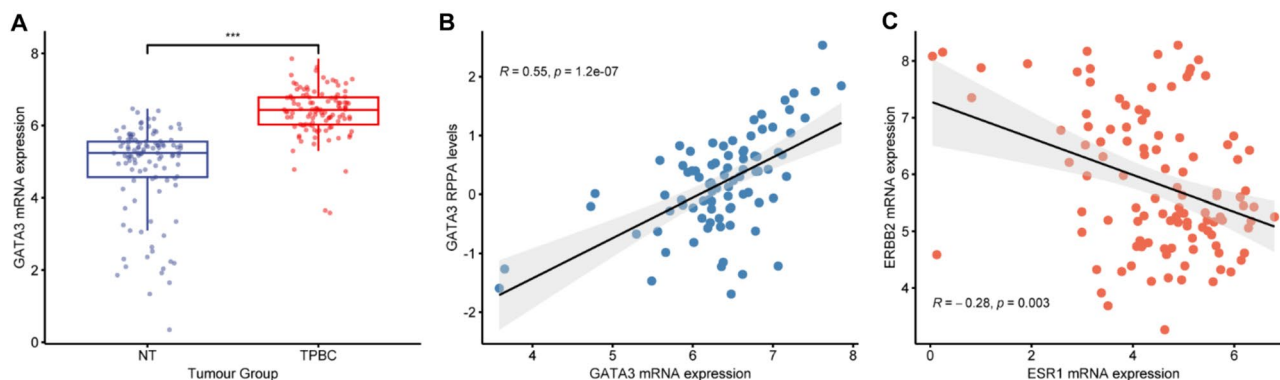


Figure 2. Expression of GATA3 and relationship between ER and HER2 in TPBC. (A) GATA3 expression difference between triple-positive breast cancer (TPBC) and normal tissue in The Cancer Genome Atlas (TCGA) database. (B) Correlation between GATA3 RPPA and mRNA expression of TPBC cases in TCGA database. (C) Correlation between ER and HER2 expression of TPBC cases in TCGA database. *** P -value < 0.001 .

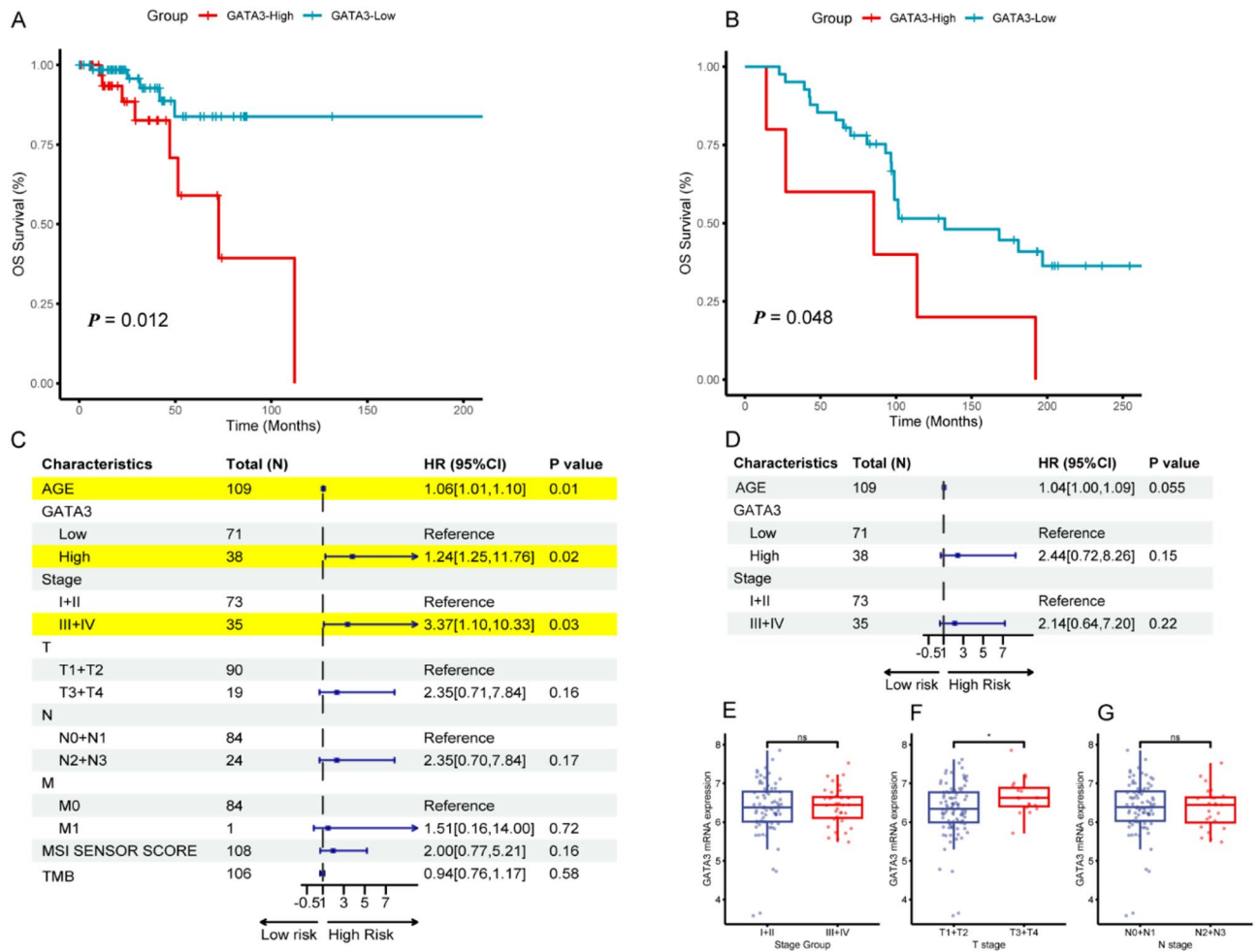


Figure 3. Prognostic value of GATA3 expression in TPBC. Kaplan–Meier curves show that the high GATA3 expression of TPBC cases in the TCGA database (A) and METABRIC dataset (B). Forest plot showing the univariate (C) and multivariate (D) Cox regression analysis results in TPBC cases. (E–G) GATA3 is concerned with clinical characteristics involving the clinical, T, and N stages. * P -value < 0.05 and NS (not significant).

our further investigation revealed that the prognostic significance of GATA3 was inconsistent across different molecular subtypes of breast cancer (Figure S1).

Enrichment analysis of GATA3 gene in TPBC

Given the close relationship between different expression levels and clinical prognosis, further investigation into the potential function of GATA3 in TPBC was conducted through GSEA analysis using the TCGA database. Immune-related gene sets were positively enriched in the GATA3 low-expression group (Fig. 4A). The high-expression group also showed a strong association with DNA mismatch repair-related gene sets (Fig. 4B). No significant differences in MSI and TMB were observed between the two groups (Figs. 4C–D). However, the mRNA expression of specific mismatch repair-related genes was notably higher in the GATA3 high-expression group (Fig. 4E).

Correlation between GATA3 and immune infiltrates

Enrichment analysis revealed significant differences in immune infiltration between the GATA3 low and high expression groups, prompting further investigation into the tumor microenvironment (TME). The TME is widely acknowledged as a crucial factor in regulating the progression of malignant tumors and influencing their response to therapies. Our findings showed that the stromal, immune, and ESTIMATE scores were higher in the GATA3 low expression group (Fig. 5A–C). Within the TME, immune cells and immune-related genes play crucial roles in maintaining homeostasis and influencing the dynamic development of the microenvironment. Consequently, we explored the impact of GATA3 expression on immune cells, revealing a significant association between low GATA3 expression and increased infiltration of lymphocytes (Fig. 5D), which was also verified by the METABRIC database (Figure S3). Furthermore, comprehensive GSVA analysis showed a negative correlation between GATA3 expression and immune-related gene set expression, especially in the T-cell receptor (TCR) signaling pathway (Fig. 5E–F). This suggests the potential involvement of GATA3 in modulating immune processes. In

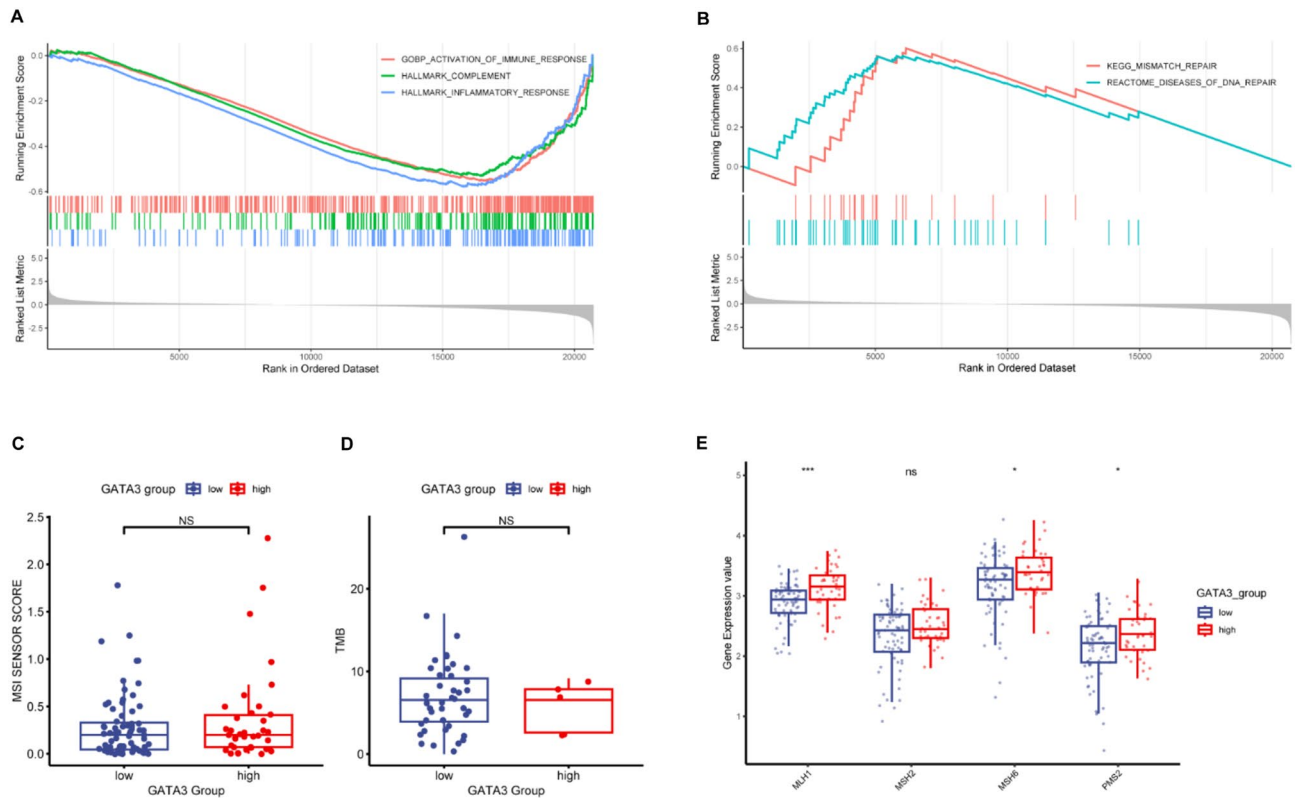


Figure 4. Enrichment analysis of the GATA3 gene in TPBC and the differences between MSI and TMB among different GATA3 expression groups. (A, B) GSEA showed the enriched pathways and gene sets in the low and high GATA3 groups. (C, D) MSI score and TMB level in the low and high GATA3 groups. (E) MMR-related genes (MLH1, MSH2, MSH6, PMS2) in the low and high GATA3 groups. * P -value < 0.05 , *** P -value < 0.001 , and NS (not significant).

addition, the low GATA3 expression group displayed higher levels of immunotherapy-related genes, such as PD-1 (PDCD1), PD-L1 (CD274), CTLA-4, and LAG3 (Fig. 5G). The IPS of PD1 (+) was significantly higher in the low-GATA3 expression group than in the high-expression group (Fig. 5H–I). These findings suggest that GATA3 could play a key role in shaping the immune environment of TPBC, potentially influencing the tumor's response to immunotherapy. Finally, oncoPredict analysis showed no significant differences in endocrine therapy responsiveness between the groups (Figure S2).

The GATA3 could be an effective predictor of NACT response in TPBC patients

The GATA3 expression of patients with pCR was significantly lower than that of patients with RD (Fig. 6A). Further investigation revealed a high expression of activated CD4⁺ T cells in the pCR group (Fig. 6B). Additionally, a statistically significant negative correlation was observed between GATA3 expression and activated CD4⁺ T cells (Fig. 6C) (Spearman: $R = -0.61$, $P < 0.05$). These findings suggest that high GATA3 expression is associated with treatment resistance and low infiltration of T lymphocytes.

Mutation and methylation analysis in different GATA3 expression groups

Genomic mutations and methylation often influence gene expression during tumor development and progression. Our investigation identified high mutation rates in the PIK3CA and TP53 genes within TPBC (Fig. 7A). Subsequently, we examined the mutation profiles across different GATA3 expression groups in TPBC. The findings indicated that the high GATA3 expression group exhibited a higher GATA3 mutation rate (Fig. 7B). Further exploration of the relationship between GATA3 methylation and mRNA expression in TPBC showed a statistically significant negative correlation (Spearman: $R = -0.46$, $P < 0.05$) (Fig. 7C). The above analysis revealed possible reasons for the differential mRNA expression of GATA3.

The relationship between TILs and GATA3 expression was validated by clinical samples

To explore the association between GATA3 protein expression and TILs, we first used immunohistochemical staining and H-score analysis to assess GATA3 expression levels. Some example images of the GATA3 IHC results were shown in Figure S4. Based on staining intensity and extent, samples were classified into the GATA3 high-expression group (Fig. 8A) and the GATA3 low-expression group (Fig. 8C). Concurrently, the same HE slides were used to assess the extent of TILs in both groups. The high GATA3 expression group exhibited fewer TILs (Fig. 8B), whereas the low GATA3 expression group displayed more TILs (Fig. 8D). Subsequent analysis

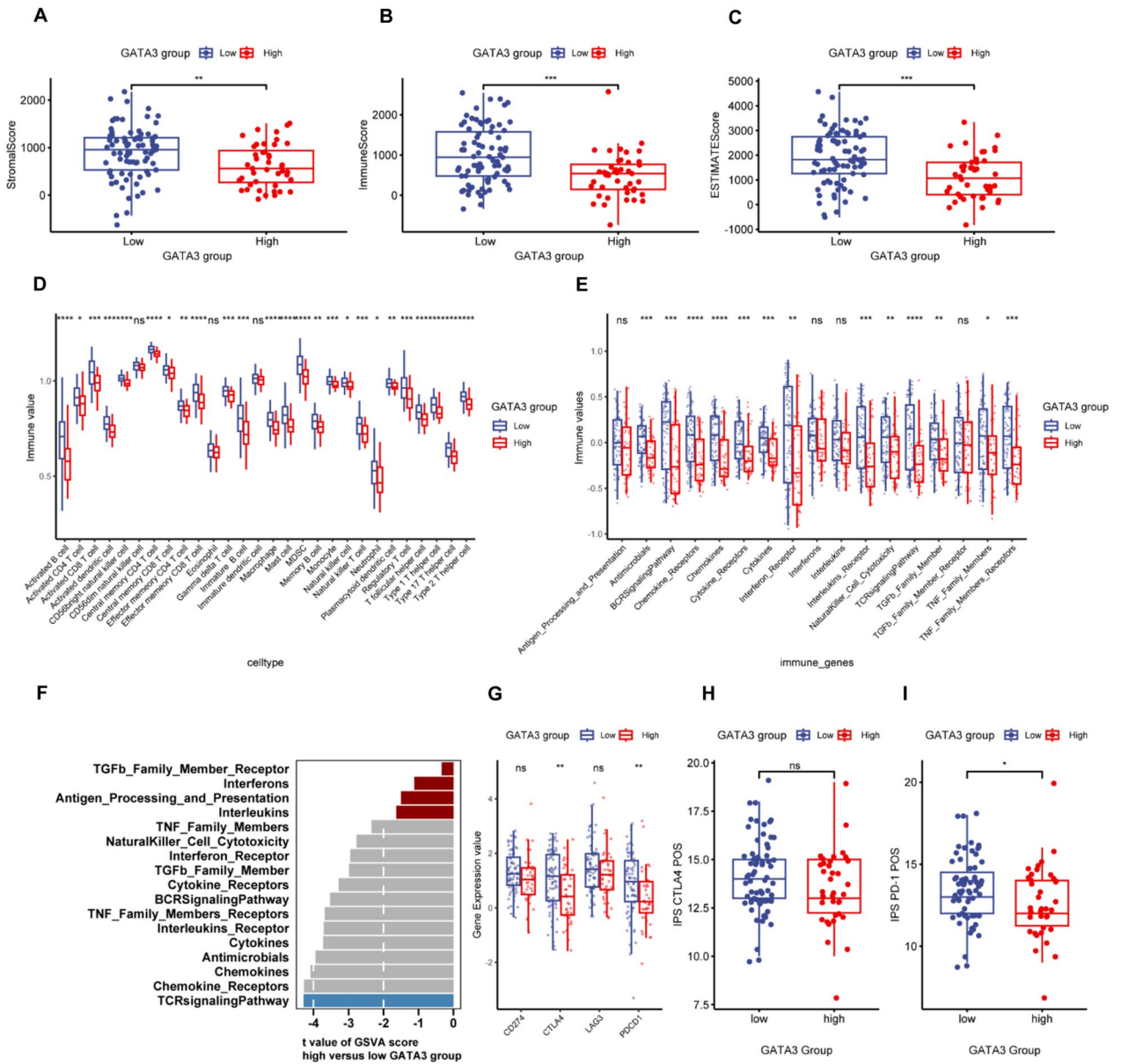


Figure 5. Association between GATA3 and immune infiltration in TPBC. (A–C) Comparison of stromal, immune, and estimate scores in the different GATA3 expression groups. (D) The infiltration ratio of 28 immune cells in the different GATA3 expression groups by the ssGSEA. (E, F) Comparison of 17 immune gene sets in the different GATA3 expression groups by the GSEA. (G) Expression of immunotherapy-related genes in the different GATA3 expression groups. (H, I) The immunophenotype scores in the different GATA3 expression groups. **P*-value < 0.05, ***P*-value < 0.01, ****P*-value < 0.001, *****P*-value < 0.0001, and NS (not significant).

revealed significantly more TILs in the low GATA3 expression group (*P* < 0.05) (Fig. 8E). Furthermore, the correlation between GATA3 expression and the clinical features of TPBC was investigated. Our findings indicated a significant association between GATA3 expression and the clinical stage of TPBC (*P* < 0.05) (Fig. 8F). However, no significant correlations were observed between GATA3 expression and age, Ki67 index, histological grade, T stage, or N stage (*P* > 0.05) (Figure S5).

Discussion

There are complex interactions between signaling pathways that drive tumor progression and response to treatment in TPBC²⁶. An important finding is a significant negative correlation between the HER2 and ER protein expression in TPBC. Specifically, inhibition of the HER pathway can lead to a compensatory upregulation of ER expression, resulting in the reactivation of ER signaling pathways. This reactivation is believed to contribute to resistance against anti-HER2 therapy²⁷. Additionally, a high frequency of PIK3CA mutations in TPBC has been implicated in mediating resistance to endocrine therapy²⁸. Due to the complex molecular alterations in TPBC

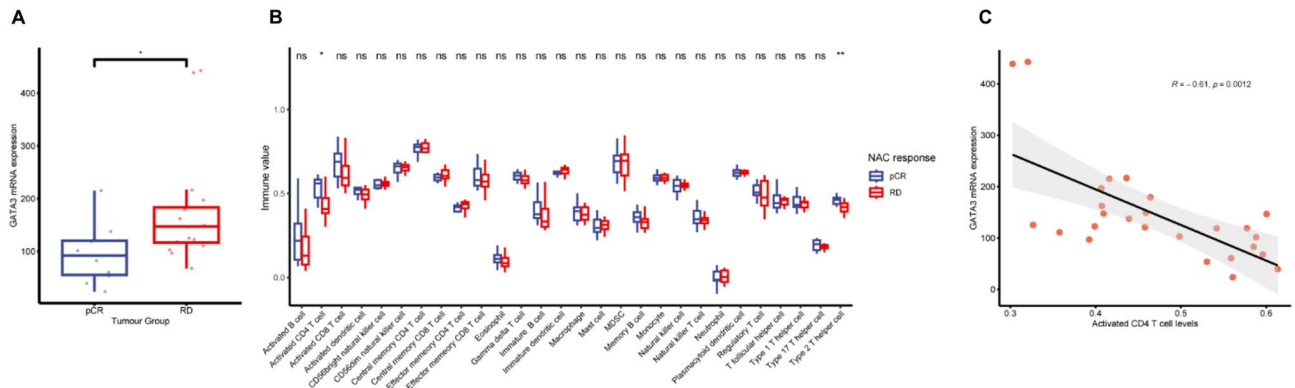


Figure 6. Association between GATA3 expression and NACT response. (A) The grouping diagram showed different GATA3 expression between pPCR and PD patients in the GSE123845 cohort. (B) The infiltration ratio of 28 immune cells between pPCR and PD patients by the ssGSEA in the GSE123845 cohort. (C) Correlation between GATA3 expression and Activated CD4⁺ T cell levels in the GSE123845 cohort. **P*-value < 0.05, ***P*-value < 0.01, and NS (not significant).

that result in varied treatment responses, further study is essential to identify reliable biomarkers for predicting treatment outcomes and guiding personalized treatment strategies.

GATA3 is an important transcription factor that regulates various cell types, including CD4⁺ T lymphocytes²⁹ and mammary epithelial cells⁸. Overexpression of GATA3 is widely recognized as a tumor suppressor, inhibiting tumorigenesis and promoting cell differentiation¹⁰. In breast cancer, GATA3 is predominantly expressed in subtypes with a better prognosis, such as luminal types and low-grade tumors, correlating its expression with improved outcomes³⁰. However, our study revealed that high expression of GATA3 is associated with poor prognosis in TPBC, suggesting that its prognostic significance varies across molecular subtypes. Importantly, this association appears to be independent of tumor size and lymph node involvement, as evidenced by the lack of statistically significant differences in GATA3 expression across different T and N stages in clinical samples. These findings highlight the potential of GATA3 as an independent prognostic biomarker in TPBC, which could help guide treatment decisions and improve patient risk stratification. Meanwhile, subsequent Kaplan–Meier analysis demonstrated that the prognostic significance of GATA3 is inconsistent across different PAM50 subtypes. This observation is consistent with recent studies indicating that GATA3 has different prognostic roles in various intrinsic subtypes of breast cancer³¹. Additionally, it is crucial to acknowledge the variability in the prognostic implications of GATA3 expression among different tumor types. For instance, higher GATA3 staining has been significantly associated with improved progression-free survival in urothelial carcinoma of the upper urinary tract³², while contrasting results have been observed in high-grade serous ovarian cancer (HGSC)³³. Therefore, the prognostic significance of GATA3 remains controversial and requires further assessment and validation.

Next, GSEA analysis was employed to uncover the potential mechanisms underlying GATA3 expression in TPBC. A significant observation was the significant enrichment of MMR-related pathways in the high GATA3 expression group. However, the relationship between GATA3 expression and MMR proteins is still under investigation, and the direct evidence supporting the specific interactions is limited. Considering that MMR deficiency typically results in MSI-H or high TMB³⁴, we further investigated the difference of MSI and TMB between different groups. We found differences in the expression of MMR-related genes among different GATA3 groups. However, these differences did not result in microsatellite instability or an increase in TMB. Additionally, MSI-H or MMR-deficient events are uncommon in breast cancer, occurring in approximately 1.9% of cases³⁵. Furthermore, the role of TMB as a predictive biomarker for response to immune checkpoint inhibitors (ICIs) remains controversial³⁶. Given these findings, it is imperative to focus on other immune molecular markers to gain a more comprehensive understanding of the tumor microenvironment (TME) and guide the development of effective treatment strategies.

TME has been recognized as a crucial factor influencing the response to various therapeutic interventions, including chemotherapy, immunotherapy, and targeted therapies³⁷. Within the TME, stromal and immune cells play pivotal roles in maintaining tissue homeostasis and influencing therapeutic outcomes³⁸. Stromal and immune scores calculated by the ESTIMATE algorithm reflect the overall immune and stromal cell infiltration within the TME²¹. Our results revealed that the stromal, immune, and ESTIMATE scores were significantly lower in the high GATA3 expression group. Additionally, we found that GATA3 expression is significantly correlated with the abundance of immune cell infiltration. However, there are few reports on the relationship between GATA3 expression and immune infiltration. A recent study identified a correlation between GATA3 expression and the abundance of immune cells in bladder carcinoma¹³. Consistently, in TPBC, the ssGSEA algorithms found that the low GATA3 expression group exhibited a high level of lymphocyte infiltration, including activated B cells, CD4⁺ T cells, and CD8⁺ T cells. To validate our analyses, we performed immunohistochemical staining of GATA3 in clinical samples of TPBC. The results showed that the group with low GATA3 expression had a higher stromal lymphocyte infiltration and lower tumor stage. Furthermore, since the high abundance of TILs is associated with a favorable prognosis in breast cancer^{39,40}, this phenomenon may explain why low GATA3 expression is linked to a favorable prognosis in TPBC.

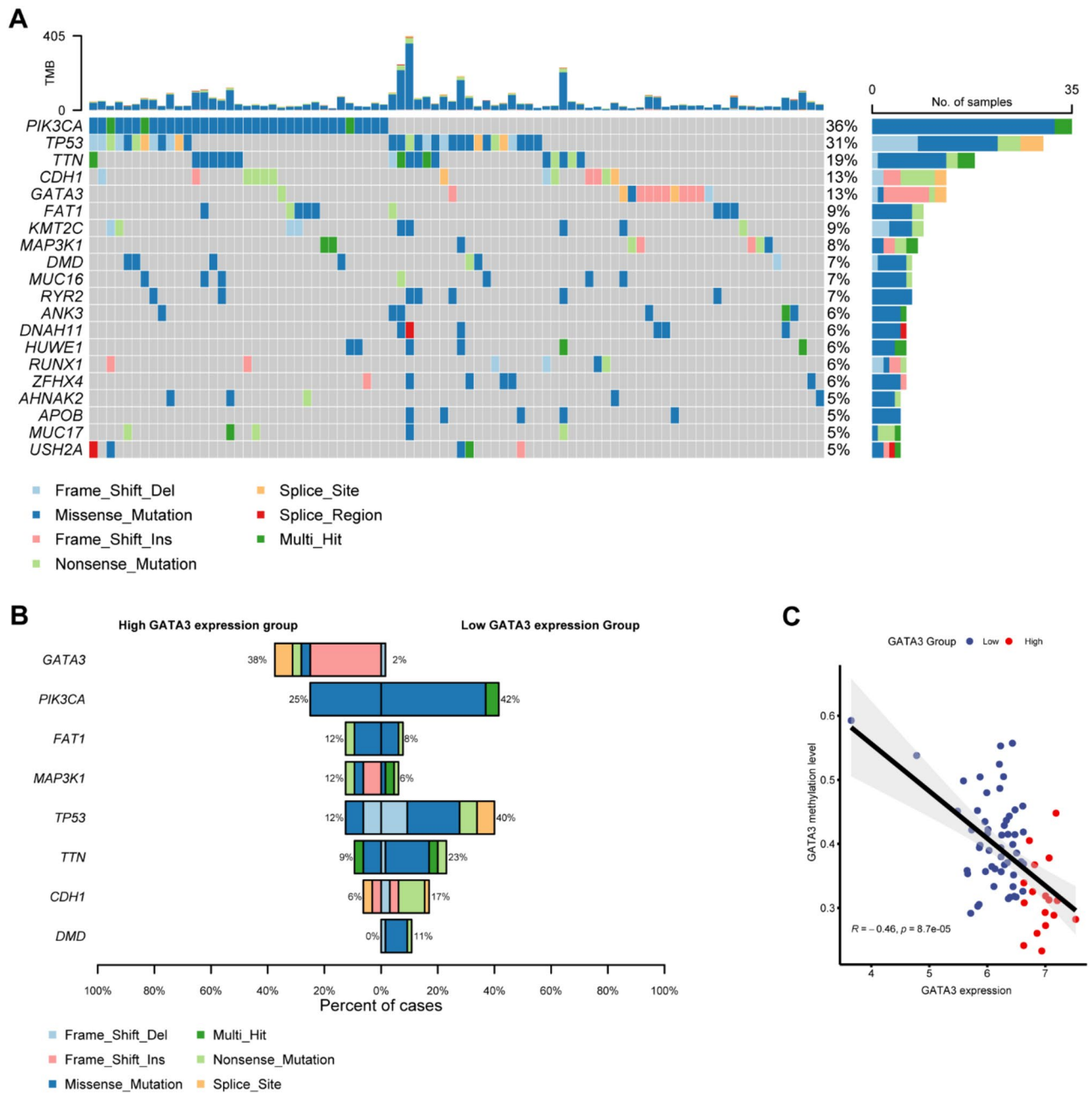


Figure 7. The Mutation and methylation analysis in the different GATA3 expression groups. (A) Waterfall plot of the top 20 mutational genes in the TPBC cases. (B) Co-bar plots of differentially mutated genes between high-GATA3 and low-GATA3 groups. (C) Correlation between GATA3 expression and GATA3 methylation in TPBC.

Due to the intersection between the HER2 and ER-mediated signaling pathways, TPBC exhibits a unique drug response and resistance mechanism distinct from other molecular subtypes. Recent clinical studies have primarily focused on the combination of HER2-targeted drug and endocrine therapy⁴¹, with less attention given to exploring the efficacy biomarker of NACT in TPBC. This study found that patients with TPBC who achieved the pCR were more likely to exhibit low expression of GATA3, suggesting that GATA3 may be related to the efficacy of NACT. Additionally, a negative relationship was observed between GATA3 expression and the levels of CD4-positive T cells. Given the close association between TILs and the outcomes of NACT, an increase in TILs is typically associated with improved chemotherapy response, higher pCR rates, and better prognosis, regardless of the molecular subtype¹². Therefore, it is reasonable to propose the following hypothesis: The expression of GATA3 could influence the efficacy of NACT for TPBC by affecting the levels of TILs. However, this hypothesis requires further validation through additional studies.

Recent immunology-related research has been actively conducted in luminal subtypes^{7,42,43}, expanding the application of immunotherapy in breast cancer. Therefore, in the face of potential HER2 inhibitors and endocrine therapy resistance in TPBC, it is essential to re-evaluate the role of immunotherapy from new perspectives.

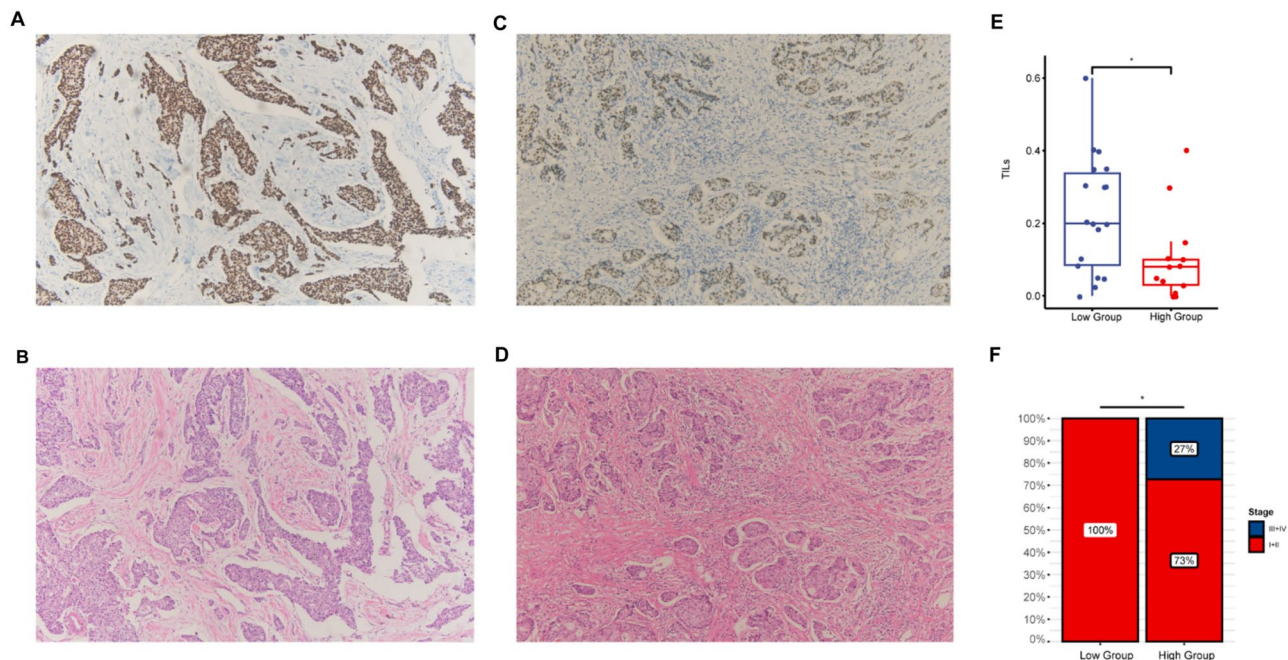


Figure 8. Validation of the expression level of GATA3 and TILs in the clinical sample. High GATA3 protein expression (A, IHC $\times 100$) but low TILs (B, HE $\times 100$) in the high-GATA3 expression group. There is low GATA3 protein expression (C, IHC $\times 100$) but high TILs (D, HE $\times 100$) in the low-GATA3 expression group. (E) The levels of TILs were higher in the low-GATA3 expression group. (F) The expression level of GATA3 correlated with the clinical stages. * P -value < 0.05 .

Regarding immunotherapy, immune checkpoint inhibitors targeting CTLA-4, PD-1, and PD-L1 have been approved by the United States Food and Drug Administration (FDA) for treating solid tumors, including breast cancer⁴⁴. Our study investigated the relationship between GATA3 expression and key immunotherapy target genes (PD-1, PD-L1, CTLA-4, and LAG3) in TPBC. The results indicate that in TPBC, lower GATA3 expression is associated with higher PD-1 levels, which could potentially enhance the effectiveness of PD-1 inhibitors, as evidenced by IPS evaluations. Multiple studies have suggested that breast cancer patients with elevated PD-1 expression are more likely to benefit from PD-1 inhibitor treatments⁴⁵. This phenomenon is frequently accompanied by increased PD-1/PD-L1 expression, correlating with high lymphocyte infiltration^{46,47}. Therefore, our findings hypothesize that GATA3 could serve as a potential immunotherapy target by regulating immune cell infiltration levels in TPBC.

Nonetheless, this study has several limitations. Primarily, the sample size was not sufficiently large, which may limit the generalizability of our findings. Additionally, the exact mechanistic roles of GATA3 in TPBC need to be validated through in vivo and in vitro studies. To reinforce the clinical significance of these results, it is imperative to conduct further research, including the collection of more TPBC cases and comprehensive retrospective studies.

Conclusions

In summary, our findings indicate that lower GATA3 expression is associated with better overall survival and a higher proportion of TILs, suggesting a potential role in enhancing the response to NACT. These results underscore the potential of GATA3 as a prognostic biomarker and therapeutic target in TPBC. Future research should focus on validating these findings in larger cohorts and exploring the underlying mechanisms through experimental studies, which could pave the way for improved therapeutic strategies in TPBC treatment.

Data availability

The TCGA dataset was obtained from <https://portal.gdc.cancer.gov/>. The METABRIC dataset was obtained from the https://www.cbioportal.org/study/summary?id=brca_metabric. The GSE123845 dataset was obtained from the <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123845>. Further inquiries can be directed to the corresponding author.

Received: 12 March 2024; Accepted: 29 July 2024

Published online: 01 August 2024

References

- Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10869–10874 (2001).

3. Elsans, D. A., Masoud, E. M., Kamel, N. A. M. H. & Ahmed, A. M. Immunohistochemical signaling pathways of triple negative and triple positive breast cancers: What is new?. *Ann. Diagn. Pathol.* **55**, 151831 (2021).
4. Geng, A., Xiao, J., Dong, B. & Yuan, S. Analysis of prognostic factors and construction of prognostic models for triple-positive breast cancer. *Front. Oncol.* **13**, 1071076 (2023).
5. Arciero, C. A. *et al.* ER+/HER2+ breast cancer has different metastatic patterns and better survival than ER-/HER2+ breast cancer. *Clin. Breast Cancer* **19**, 236–245 (2019).
6. Geyer, F. C., Rodrigues, D. N., Weigelt, B. & Reis-Filho, J. S. Molecular classification of estrogen receptor-positive/luminal breast cancers. *Adv. Anat. Pathol.* **19**, 39–53 (2012).
7. Dieci, M. V. *et al.* Neoadjuvant Chemotherapy and Immunotherapy in Luminal B-like Breast Cancer: Results of the phase II GIADA Trial. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **28**, 308–317 (2022).
8. Kouros-Mehr, H., Kim, J., Bechis, S. K. & Werb, Z. GATA-3 and the regulation of the mammary luminal cell fate. *Curr. Opin. Cell Biol.* **20**, 164–170 (2008).
9. Van de Walle, I. *et al.* GATA3 induces human T-cell commitment by restraining Notch activity and repressing NK-cell fate. *Nat. Commun.* **7**, 11171 (2016).
10. Bc, M. *et al.* GATA3 expression in advanced breast cancer: prognostic value and organ-specific relapse. *Am. J. Clin. Pathol.* (2015).
11. Jacquemier, J. *et al.* Association of GATA3, P53, Ki67 status and vascular peritumoral invasion are strongly prognostic in luminal breast cancer. *Breast Cancer Res. BCR* **11**, R23 (2009).
12. Denkert, C. *et al.* Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol.* **19**, 40–50 (2018).
13. Zhang, Q. *et al.* GATA3 predicts the tumor microenvironment phenotypes and molecular subtypes for bladder carcinoma. *Front. Surg.* **9**, 860663 (2022).
14. Colaprico, A. *et al.* TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
15. Silva, T. C. *et al.* TCGA workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. *F1000Research* **5**, 1542 (2016).
16. Mounir, M. *et al.* New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. *PLoS Comput. Biol.* **15**, e1006701 (2019).
17. Allison, K. H. *et al.* Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **38**, 1346–1366 (2020).
18. Park, Y. H. *et al.* Chemotherapy induces dynamic immune responses in breast cancers that impact treatment outcome. *Nat. Commun.* **11**, 6175 (2020).
19. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov. Camb. Mass* **2**, 100141 (2021).
20. Hänzelmann, S., Castelo, R. & Guinney, J. GSEA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinf.* **14**, 7 (2013).
21. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
22. Charoentong, P. *et al.* Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* **18**, 248–262 (2017).
23. Maeser, D., Gruener, R. F. & Huang, R. S. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief. Bioinform.* **22**, bbab260 (2021).
24. Kim, S. *et al.* Expression patterns of GATA3 and the androgen receptor are strongly correlated in patients with triple-negative breast cancer. *Hum. Pathol.* **55**, 190–195 (2016).
25. Salgado, R. *et al.* The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: Recommendations by an international TILs working group 2014. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **26**, 259–271 (2015).
26. Kay, C. *et al.* Current trends in the treatment of HR+/HER2+ breast cancer. *Future Oncol. Lond. Engl.* **17**, 1665–1681 (2021).
27. Giuliano, M., Trivedi, M. V. & Schiff, R. Bidirectional crosstalk between the estrogen receptor and human epidermal growth factor receptor 2 signaling pathways in breast cancer: Molecular basis and clinical implications. *Breast Care Basel Switz.* **8**, 256–262 (2013).
28. Guarneri, V. *et al.* PIK3CA mutation in the ShortHER randomized adjuvant trial for patients with early HER2+ breast cancer: Association with prognosis and integration with PAM50 subtype. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **26**, 5843–5851 (2020).
29. Yagi, R., Zhu, J. & Paul, W. E. An updated view on transcription factor GATA3-mediated regulation of Th1 and Th2 cell differentiation. *Int. Immunol.* **23**, 415–420 (2011).
30. Fararjeh, A.-F.S. *et al.* The impact of the effectiveness of GATA3 as a prognostic factor in breast cancer. *Hum. Pathol.* **80**, 219–230 (2018).
31. Querzoli, P. *et al.* GATA3 as an adjunct prognostic factor in breast cancer patients with less aggressive disease: A study with a review of the literature. *Diagnostics.* (2021).
32. Inoue, S. *et al.* GATA3 immunohistochemistry in urothelial carcinoma of the upper urinary tract as a urothelial marker and a prognosticator. *Hum. Pathol.* **64**, 83–90 (2017).
33. Chen, H. J. *et al.* GATA3 as a master regulator and therapeutic target in ovarian high-grade serous carcinoma stem cells: GSKJ4 targeting GATA3/UTX interaction eliminate HGSC stem cells. *Int. J. Cancer* **143**, 3106–3119 (2018).
34. Jiang, M. *et al.* Alterations of DNA damage response pathway: Biomarker and therapeutic strategy for cancer immunotherapy. *Acta Pharm. Sin. B* **11**, 2983–2994 (2021).
35. Cheng, A. S. *et al.* Mismatch repair protein loss in breast cancer: clinicopathological associations in a large British Columbia cohort. *Breast Cancer Res. Treat.* **179**, 3–10 (2020).
36. de Melo Gagliato, D., Buzaid, A. C., Perez-Garcia, J. & Cortes, J. Immunotherapy in breast cancer: Current practice and clinical challenges. *BioDrugs Clin. Immunother. Biopharm. Gene Ther.* **34**, 611–623 (2020).
37. Xiao, Y. & Yu, D. Tumor microenvironment as a therapeutic target in cancer. *Pharmacol. Ther.* **221**, 107753 (2021).
38. Mao, X. *et al.* Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. *Mol. Cancer* **20**, 131 (2021).
39. Dieci, M. V. *et al.* Prognostic and predictive value of tumor-infiltrating lymphocytes in two phase III randomized adjuvant breast cancer trials. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **26**, 1698–1704 (2015).
40. Mao, Y. *et al.* The prognostic value of tumor-infiltrating lymphocytes in breast cancer: A systematic review and meta-analysis. *PLoS One* **11**, e0152500 (2016).
41. Niu, N. *et al.* A multicentre single arm phase 2 trial of neoadjuvant pyrotinib and letrozole plus dalpiciclib for triple-positive breast cancer. *Nat. Commun.* **13**, 7043 (2022).
42. Goldberg, J. *et al.* Estrogen receptor mutations as novel targets for immunotherapy in metastatic estrogen receptor-positive breast cancer. *Cancer Res. Commun.* **4**, 496–504 (2024).
43. Lee, S. *et al.* B-cell-mediated immunity predicts survival of patients with estrogen receptor-positive breast cancer. *JCO Precis. Oncol.* **8**, e2300263 (2024).
44. Barzaman, K. *et al.* Breast cancer immunotherapy: Current and novel approaches. *Int. Immunopharmacol.* **98**, 107886 (2021).

45. Li, C. J., Lin, L. T., Hou, M. F. & Chu, P. Y. PD-L1/PD-1 blockade in breast cancer: The immunotherapy era (Review). *Oncol. Rep.* **45**, 5–12 (2021).
46. Kitano, A. *et al.* Tumour-infiltrating lymphocytes are correlated with higher expression levels of PD-1 and PD-L1 in early breast cancer. *ESMO Open* **2**, e000150 (2017).
47. Wimberly, H. *et al.* PD-L1 expression correlates with tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy in breast cancer. *Cancer Immunol. Res.* **3**, 326–332 (2015).

Author contributions

XC designed the study. Both XC and WZ were involved in the collection and analysis of data. XC drafted this manuscript. XC, WZ, and SL interpreted the data. YH and XT reviewed the manuscript. QY supervised the study. All authors have approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-68788-8>.

Correspondence and requests for materials should be addressed to X.T. or Q.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024